A novel segmentation framework for uveal melanoma in magnetic resonance imaging based on class activation maps

Huu-Giao Nguyen[1,2,3] HUU.NGUYEN@ARTORG.UNIBE.CH
[1] *Proton Therapy Center, Paul Scherrer Institut, ETH Domain, Villigen, Switzerland*
[2] *Ophthalmic Technology Lab., ARTORG Center, University of Bern, Switzerland*
[3] *Radiology Department, Lausanne University Hospital (CHUV), Switzerland*

Alessia Pica[1] and **Jan Hrbacek**[1] and **Damien C. Weber**[1,4]
FIRSTNAME.LASTNAME@PSI.CH
[4] *Radiation Oncology Department, Inselspital, University of Bern, Switzerland*
Francesco La Rosa[5] FRANCESCO.LAROSA@EPFL.CH
[5] *Signal Processing Lab., Ecole Polytechnique Fédérale de Lausanne, Switzerland*

Ann Schalenbourg[6] ANN.SCHALENBOURG@FA2.CH
[6] *Adult Ocular Oncology Unit, Jules-Gonin Eye hospital, Lausanne, Switzerland*

Raphael Sznitman[2] RAPHAEL.SZNITMAN@ARTORG.UNIBE.CH
Meritxell Bach Cuadra[3,5,7] MERITXELL.BACHCUADRA@UNIL.CH
[7] *Medical Image Analysis Laboratory, CIBM, University of Lausanne, Switzerland*

Abstract

An automatic and accurate eye tumor segmentation from Magnetic Resonance images (MRI) could have a great clinical contribution for the purpose of diagnosis and treatment planning of intra-ocular cancer. For instance, the characterization of uveal melanoma (UM) tumors would allow the integration of 3D information for the radiotherapy and would also support further radiomics studies. In this work, we tackle two major challenges of UM segmentation: 1) the high heterogeneity of tumor characterization in respect to location, size and appearance and, 2) the difficulty in obtaining ground-truth delineations of medical experts for training. We propose a thorough segmentation pipeline consisting of a combination of two Convolutional Neural Networks (CNN). First, we consider the class activation maps (CAM) output from a Resnet classification model and the combination of Dense Conditional Random Field (CRF) with a prior information of sclera and lens from an Active Shape Model (ASM) to automatically extract the tumor location for all MRIs. Then, these immediate results will be inputted into a 2D-Unet CNN whereby using four encoder and decoder layers to produce the tumor segmentation. A clinical data set of 1.5T T1-w and T2-w images of 28 healthy eyes and 24 UM patients is used for validation. We show experimentally in two different MRI sequences that our weakly 2D-Unet approach outperforms previous state-of-the-art methods for tumor segmentation and that it achieves equivalent accuracy as when manual labels are used for training. These results are promising for further large-scale analysis and for introducing 3D ocular tumor information in the therapy planning.

Keywords: Activation map, CAM, Unet, tumor segmentation, Uveal melanoma

1. Introduction

UM is the most common primary intraocular malignancy in the white adult population, making up 79-88% of primary intraocular cancers (Singh et al., 2014; Lemke et al., 1999). Several 2-

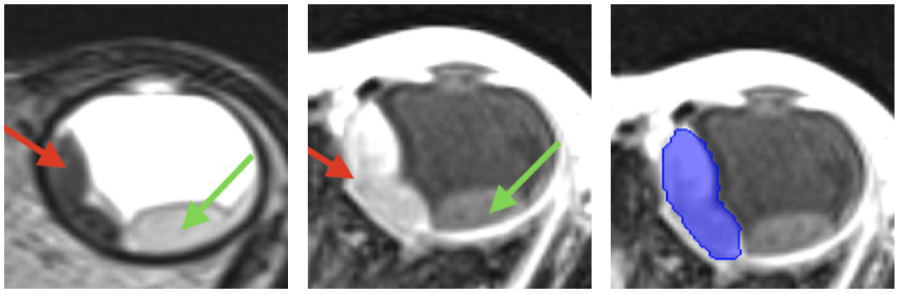


Figure 1: Example of UM in MRI: (left) T2-w; (center) T1-w; (right) manual tumor segmentation. Red & green arrows indicate the tumor & retinal detachment respectively.

dimensional and 3-dimensional imaging modalities such as 2D Fundus imaging and 3D computed tomography are needed to properly characterize the tumor, its growth and for any follow-up care. Recently, 3D MRI is raising interest in the treatment of ocular tumors (de Graaf et al., 2014; Tartaglione et al., 2014). Thanks to its high spatial resolutions and high intrinsic contrast, 3D MRI allows for clear overall improved discrimination between anatomical structures and different pathological regions (Tartaglione et al., 2014) (see Fig. 1). An automatic extraction of quantitative and reliable information of ocular tumors in MR images, i.e the location, size, texture, morphology and distribution of pathological tissues, would be a breakthrough in current diagnosis, follow-up and therapy planning procedures. Ultimately, having 3D patient-specific eye model provide an optimal solution for radiation therapy in the framework of personalized medicine in order to plan and deliver very conformal radiation dose to to the tumor while minimizing irradiation of critical structures (Beenakker et al., 2015).

Few automated methods have been tailored for ocular tumors segmentation from MRI. First attempts were dedicated to the segmentation of retinoblastoma in children. Two deep learning techniques were proposed, a 3D-Unet (Nguyen et al., 2018a) and a 3D-CNN (Ciller et al., 2017), based on rather small datasets of 16 and 32 retinoblastoma eyes. The tumor segmentation performance reported in those pioneer works was , however, relatively low, with an average Dice similarity coefficient (DSC) measurement of around 62%. Recently, UM tumors have been tackled in (Hassan et al., 2018), based on image registration and threshold of MRI, though only four cases were qualitatively evaluated. One of the major limitations of these approaches, affecting specifically supervised techniques, is the lack of manual delineations. Actually, we think the above 3D deep learning architectures were highly limited by the low number of training samples available. Unfortunately, having such input labels is very tedious, time consuming and not easily available in practice.

Weakly supervised methods based on CAM for the segmentation of pathological tissues have recently received a great attention, e.g. pulmonary nodules in CT (Feng et al., 2017) or diabetic retinopathy lesions in retinal fundus images (Gondal et al., 2017). Here, our first aim is to present an ocular tumor segmentation framework without the need of manual annotations for training. To this end, we propose an end-to-end tumor segmentation framework with two CNNs for 2D images extracted from 3D volume MRI. Our approach is based on the estimation of CAMs from a CNN architecture that classifies whether there is a tumor or not in the image. Afterwards, we refine the CAMs by combining an ASM segmentation of the eye structures (Nguyen et al., 2018b) with a dense

Table 1: MR imaging acquisition parameters at 1.5T with a surface coil.

	Repetition time(ms)	Echo time (ms)	Flip Angle	Voxel size (mm^3)	FOV (Voxels)	Healthy	UM
T1-VIBE	6.55	2.39	12°	0.5x0.5x0.5	256x256x80	28 eyes	24 eyes
T2-SPACE	1400	185	150°	0.5x0.5x0.5 and 0.82x0.82x0.8	256x256x80	25 eyes	22 eyes

CRF to maximize label agreement between similar pixels in images. Finally, we use these refined CAMs as input training data for a 2D-Unet segmentation (Ronneberger et al., 2015). The proposed framework is cheaper in training data (only sclera segmentations are needed for the ASM) and outperforms in segmentation compared existing deep learning approaches (Nguyen et al., 2018a; Rosa et al., 2018).

A second major contribution of this work is the quantitative evaluation of several 2D and 3D architectures for the UM segmentation. To the best of our knowledge this is the first study reporting automated segmentation accuracy for such ocular tumor. Our proposed segmentation technique will be compared with previous related work: 1) our previous work tailored for retinoblastoma tumors in children and based on a 3D-Unet (Nguyen et al., 2018a), with a 2D-Unet using manual labels as training from an expert, and 2) a cascade of two 3D patch-wise CNNs used for lesion segmentation in Multiple Sclerosis (Rosa et al., 2018).

2. Dataset

MR acquisitions were performed by a 1.5T Siemens scanner with surface coil for both T1w and T2w contrasts at the Paul Scherrer Institute. A set of 16 healthy volunteers (mean age 29 ± 5.4 y.o., range $[23-46]$ years) and 24 UM patients (mean age 63 ± 14 y.o., range $[36-74]$ years) was considered. The cohort median eye size was 24.4mm of diameter (range, 22.1-26.5). Tab. 1 shows the different parameters used for the MRI acquisition protocol. The study was approved by the Ethics Committee of the involved institutions and all subjects (anonymized and de-identified) provided written informed consent prior to participation.

Images were pre-processed as follows. First, an anisotropic diffusion filtering (Perona and Malik, 1990) was applied to reduce noise without removing significant image content. Second, we applied the N4 algorithm (Tustison et al., 2010) to correct for bias field variations and performed histogram-based intensity normalization (Nyul et al., 2000) for an intensity normalisation. Finally, in order to improve the performance in segmentation and computation time, the whole MRI was cropped using a volume of interest of 64x64x64 voxels centered in the eye.

Manual delineations were done by radiation oncologist expert for 16 UM patients and all healthy eyes using Velocity software(Varian Medical System, Palo Alto, CA). First, segmentation for sclera, lens and tumor was done individually through intensity threshold. Second, manual editing was performed to refine borders and remove outlier regions.

3. Proposed segmentation framework

The proposed framework is over-viewed in Fig. 2. It mainly consists of the concatenation of a 2D ResNet model (He et al., 2016) to classify MRI slices (with or without tumor) that combined with

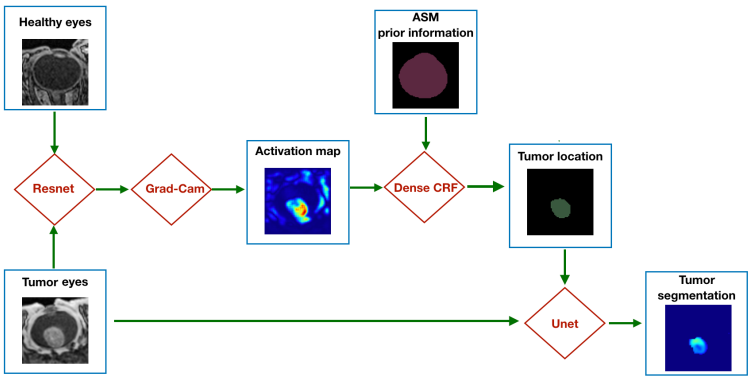
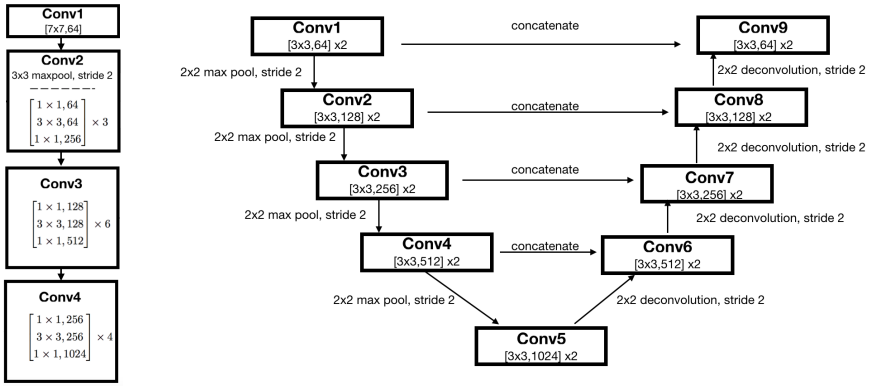


Figure 2: Main pipeline of our approach proposed.



(*a*) Resnet (*b*) 2DUnet

Figure 3: Neural network architectures used.

the ASM (Nguyen et al., 2018b) and a CRF (Krähenbühl and Koltun, 2011) will provide the initial labels for training a 2D-Unet (Ronneberger et al., 2015) model to segment tumor.

ResNet classification. In this work, we used the ResNet model (He et al., 2016) for classification of 2D input images with the score presence or absence of tumor. ResNet has the advantage of avoiding the degradation problem of deep CNN, which occurs when the accuracy gets saturated and rapidly degrades as a result of an increasing network depth. The ResNet replaces a direct mapping of input x to its score y with a function $F(x)$ by a residual function using $F(x)+x$, where $F(x)$ and x represents the stacked non-linear layers and the identity function respectively. The architecture of our ResNet is in Fig. 3(a).

Tumor location by CAM. Considering a CNN-based classification, each layer retains detailed spatial information of object and its characterization used by network to identify the category. CAMs (Zhou et al., 2016) produce such class-discriminative localization using a linear combination of $f_k(i,j)$ represent the activation of unit k in the last convolutional layer at spatial location (i,j) and the weight w_k^c corresponding to class c for unit k:

$$M^c(i,j) = \sum_k w_k^c f_k(i,j) \tag{1}$$

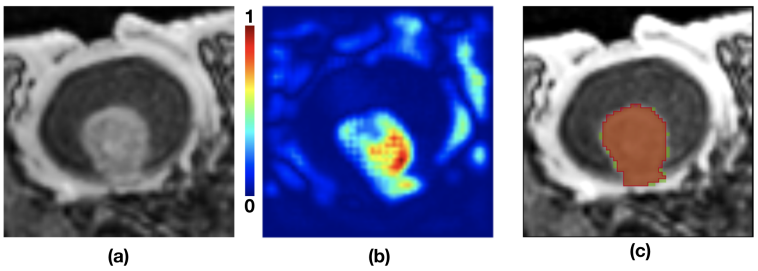


Figure 4: Example of tumor location: (a) original image; (b) grad-cam result; (c) ASM constrain for sclera (in red), DenseCRF result (in green - Dice overlap 95%).

As a generalization of CAM, Grad-CAM heat map (Selvaraju et al., 2017) is constructed from the liner combination of the importance weights α_k^c and feature maps A of a convolutional layer with respect to the gradient of the score y^c of class c:

$$L^c = ReLU\left(\sum_k \alpha_k^c A_k\right) \tag{2}$$

Refinement. We apply a dense CRF (Krähenbühl and Koltun, 2011) to maximize label agreement between similar pixels. The Dense CRF incorporates unary potentials of individual pixels and pair-wise potentials (in terms of appearance and smoothness) on neighboring pixels to provide more homogeneous regions. Considering as input image the 2D MRI slice I (either T1w or T2w) and a probability map P provided by the Grad-CAM, the unary potential is defined to be the negative log-likelihood $\psi_u(z_i) = -logP(z_i|I)$, where z_i the predicted label of voxel i. The pair-wise potential has the form $\psi_p(z_i,z_j) = \mu(z_i,z_j)k(f_i,f_j)$, where μ is a label compatibility function and $k(f_i,f_j)$ is characterized by integrating two Gaussian kernels of appearance (first term) and smoothness (second term), as follows:

$$k(f_i,f_j) = w_1 \exp\left(-\frac{|p_i-p_j|^2}{2\theta_1^2} - \frac{|I_i-I_j|^2}{2\theta_2^2}\right) + w_2 \exp\left(-\frac{|p_i-p_j|^2}{2\theta_3^2}\right), \tag{3}$$

where p_i are pixel locations, I_i are pixel intensities, w_l are weight factor between the two terms, and the θ's are tunable parameters of the Gaussian kernels. The Gibbs energy of CRF model is then given by $\sum(\psi_u(z_i),\psi_p(z_i,z_j))$. Here, we apply the inference of Dense CRF with different iterative numbers $\{5,20,50\}$ where the mean field approximation is computed by minimizing the KL-divergence while constraining the distributions.

Finally, prior information about the healthy structures such as the sclera and lens was used as tumor location constraint. Our previous work (Nguyen et al., 2018b) evaluated the DSC values of the sclera ($94.5\% \pm 1.6$) and lens ($88.3\% \pm 2.8$) on the same data set. The ASM segmentation is used to constrain the result of the CRF as shown in Fig. 4.

UNet. Similar to the original UNet method (Ronneberger et al., 2015), we consider an encoder and decoder network that takes as input 2D image with tumor and label output of Grad-cam. Each encoding pathway contains 4 layers that effectively changes the feature dimension (i.e. 64, 128, 256, 512,1024 - Fig. 3(b)). The same architecture accounts for the decoding pathway. In each case, two

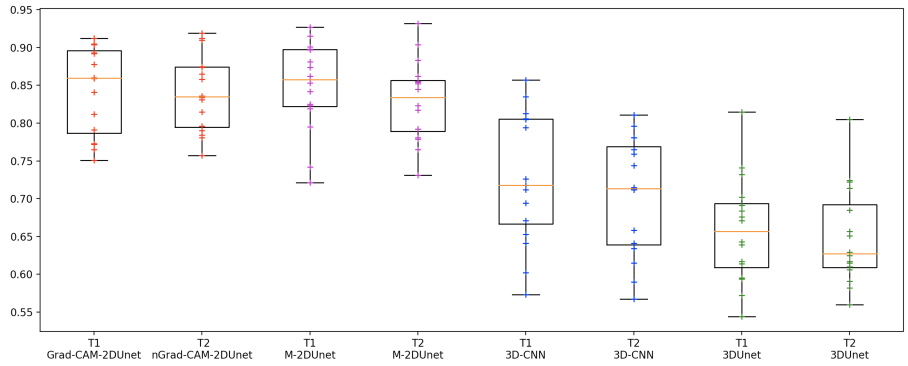


Figure 5: Boxplot on DSC of tumors segmentation on T1-w and T2-w MR images.

Wilcoxon test compared with	M-2DUnet	3D-CNN	3D-Unet
p_value on T1	0.679	0.00320	0.000437
p_value on T2	0.147	0.00097	0.000436

Table 2: Wilcoxon signed rank test on DSC comparing our method with the other strategies.

convolution layers 3x3 are used with rectified linear unit (Relu) operations and with zero padding. Between two layers in the encoder pathway, 2x2 max pooling with strides of two in each dimension are used. In the decoder pathway, a 2x2 deconvolution layer with strides of two is firstly used. Concatenation is performed to connect the output tensors of two layers of the encoder and decoder pathways at same level. To train our network, we used the Adam optimizer and the binary cross entropy loss function. Softmax is used to extract probability maps for each class. Data augmentation including rotation, shift as well as elastic deformation was applied (Simard et al., 2003).

4. Quantitative evaluation

We computed the DSC value of the predicted output as compared to the manual segmentation for the quantitative evaluation of all the automated techniques. For the 16 patients with manual segmentation, we used a leave-one-out cross-validations strategy, i.e., iteratively chose one eye as a test case, two other random eyes as validation cases while the remaining subjects are used as the training set. Moreover, to show the advantage of the proposed weakly learning 8 additional patients without manual segmentation are also included into the training set. The average number of 2D slices (containing the tumor) extracted from 3D volume of patient's eyes is 45 (range [25-60]), overall is 925 images.

Resnet binary classification model construction is trained including also tumor-free eyes. In this stage, 1915 2D slices extracted from 28 healthy eyes are also added into training set, i.e our training set have 2840 images of healthy and pathological eyes. CAMs are estimated based on all 2D images of 24 UM patient. For 2D-Unet, depending on the patients leaved out for test and validation set, around 850 2D images with tumors were selected for training.

Our framework (Grad-CAM-2DUnet) is evaluated in comparison with three baseline deep learning architectures. First, the same 2D-Unet architecture included in our framework will be used trained with the expert manual delineations instead of using the refined activation maps (M-2DUnet).

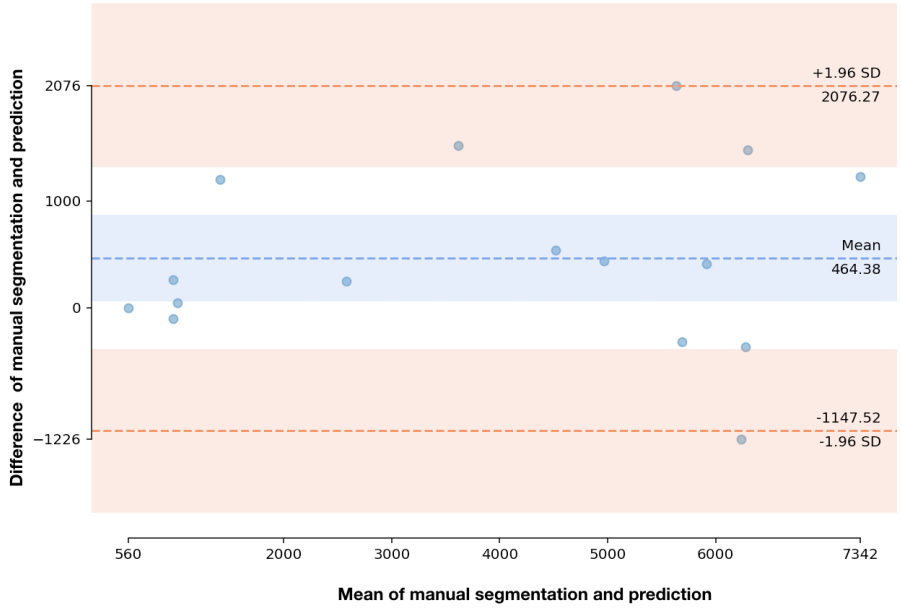


Figure 6: The Bland-Altman plot of differences (number of volume voxels) between Grad-CAM-2DUnet method's result and manual segmentation.

Second, our previous 3D-Unet (Nguyen et al., 2018a) tested on retinoblastoma patients (3D-Unet). It is composed of 4 layers of encoder and decoder pathway with different feature sizes (i. e., 32, 64, 128, 256, 512); 3x3x3 convolution with PRelu and 2x2x2 max pooling. Third, a cascade of two 3D patch-wise convolutional neural networks (Valverde et al., 2017; Rosa et al., 2018) (3D-CNN) that reported high accuracy in segmenting white matter lesions. It is composed of with 4 convolutional layers ([3x3x3,32]x2; [3x3x3,64]x2); patch-size is 11x11x11 (input images interpolated to 256x256x256).

Fig.5 shows the boxplot on DSC of four tumor segmentation methods for both T1-w and T2-w sequences, where 3DUnet with 65.8 ± 6.8 (64.9 ± 6.3) and 3D-CNN with 72.6 ± 8.2 (70.5 ± 7.5) perform in average 10% worst than the 2D-Unet strategies. This can be explained by the increased training set available in 2D as compared to the few training data in 3D. Thus, despite image acquisition is done in 3D and with a very nice isotropic spatial resolution, 2D approaches perform better. Let us note that differences in DSC were statistically significant (Wilcoxon signed rank test, $p < 0.005$) when comparing Grad-CAM-2DUnet with 3D-UNET and 3D-CNN in both T1w and T2w scenarios (Tab.2).

Our weakly supervised framework Grad-CAM-2DUnet with average Dice of 84.5 ± 5.6 for T1-w and 83.9 ± 4.9 for T2-w performs similarly to the M-2DUnet (using muanual segmentations for training) with 84.8 ± 5.7 and 82.9 ± 5.2 for T1-w and T2-w, respectively. No statistical differences were found neither for T1-w nor T2-w. The mean False positive and True positive fractions are 0.02 and 0.82 respectively when we compared our Grad-CAM-2DUnet prediction with manual segmentation. Fig.6 shows a Bland-Altman difference plot of the 3D volume comparison of manual segmentation and our prediction. This result shows the relevance of our solution for replacing the

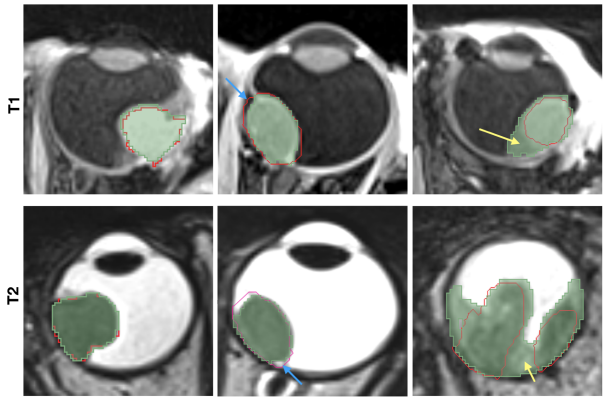


Figure 7: Qualitative results of our tumor prediction (in green) & manual segmentation (in red) on T1w and T2w: (left) high overlap; (center) automated segmentation fixed better the intensity contour (blue arrows); (right) low accuracy: our prediction cannot separate tumor and retinal detachment (yellow arrows).

costly manual annotations by free refined activation maps. Fig.7 shows qualitative result of the tumor segmentation with our proposed approach.

5. Conclusion

In this paper, we introduced an automatic and effective deep learning based approach that allows a quantitative image analysis of eye tumor tissues in adults that could further support clinicians to tailor the radiation therapy to the UM in eye tumor patients. The proposed approach takes advantage of CAMs combined with conditional random field and active shape models to provide an end-to-end segmentation without need of tumor annotations of medical experts. The paper also provides an evaluation of several 2D and 3D deep learning strategies for the UM segmentation. To our knowledge, this is the first set of techniques that have been proposed for the segmentation of UM, reporting very high accuracy in average. Our study, based on a 3D high-resolution dataset of 24 tumors, demonstrates that the best strategies for tumor segmentation make use of 2D slices instead of 3D whole volumes, that is including more data for training. Our weakly supervised framework provides a solid reliable computer-aided tool to further large-scale evaluation of ocular tumors based on MR imaging features for an enhancing a shift towards non-invasive clinical procedures.

Acknowledgments

This work is funded by the Swiss Cancer Research foundation (grant no. GAP-CRG-201602) and is supported by the Center of Biomedical Imaging of Geneva-Lausanne Universities and EPFL, the Fondation Leenaards and Fondation Louis-Jeantet. FLR is funded by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie project TRABIT (agreement No 765148).

References

J. Beenakker, D. Shamonin, A. Webb, G. Luyten, and B. Stoel. Automated retinal topographic maps measured with magnetic resonance imaging. *Investig. Ophthalmol. Vis. Sci.*, 56:1033–1039, 2015.

C. Ciller, S. De Zanet, K. Kamnitsas, P. Maeder, B. Glocker, F. Munier, D. Rueckert, J-P. Thiran, M. Bach Cuadra, and R. Sznitman. Multi-channel mri segmentation of eye structures and tumors using patient-specific features. *PLoS ONE*, 12(3), 2017.

P. de Graaf, S. Göricke, F. Rodjan, P. Galluzzi, P. Maeder, J. Castelijns, and H. Brisse. Guidelines for imaging retinoblastoma: imaging principles and mri standardization. *Pediatric radiology*, pages 2–14, 2014.

X. Feng, J. Yang, A. Laine, and E. Angelini. Discriminative localization in cnns for weakly-supervised segmen- tation of pulmonary nodules. *MICCAI*, page 568–576, 2017.

W. Gondal, J. Köhler, R. Grzeszick, G. Fink, and M. Hirsch. Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images. *arXiv preprint arXiv:1706.09634*, 2017.

M. Hassan, D. Shamonin, R. Shahzad, A. Webb, B. Stoel, and J-W. Beenakker. Automated analysis of eye tumor mr-images for an improved treatment determination. *ISMRM*, 2018.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CVPR*, pages 770–778, 2016.

P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Adv. in Neural Information Processing Systems*, pages 109–117, 2011.

A. Lemke, N. Hosten, N. Bornfeld, N. Bechrakis, A. Schüler, M. Richter, C. Stroszczynski, and R. Felix. Uveal melanoma: correlation of histopathologic and radiologic findings by using thin-section mr imaging with a surface coil. *Radiology*, 210(3):775–783, 1999.

H-G. Nguyen, A. Pica, P. Maeder, A. Schalenbourg, M. Peroni, J. Hrbacek, DC. Weber, M. Bach Cuadra, and R. Sznitman. Ocular structures segmentation from multi-sequences mri using 3d unet with fully connected crfs. *Comput. Path. & Opht. Med. Image Analysis*, pages 167–175, 2018a.

H-G. Nguyen, R. Sznitman, P. Maeder, A. Schalenbourg, M. Peroni, J. Hrbacek, DC. Weber, A. Pica, and M. Bach Cuadra. Personalized anatomic eye model from t1-weighted vibe mr imaging of patients with uveal melanoma. *Journal of Radiation Oncology, Biology, Physics*, 2018b.

L.G. Nyul, J.K. Udupa, and Xuan Zhang. New variants of a method of mri scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–50, 2000.

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 12(7):629–639, 1990.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *MICCAI*, 9351:234–241, 2015.

F. La Rosa, M. Fartaria, T. Kober, J. Richiardi, C. Granziera, J-P. Thiran, and M. Bach Cuadra. Shallow vs deep learning architectures for white matter lesion segmentation in the early stages of multiple sclerosis. *MICCAI workshop*, 2018.

R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *In arX- iv:1610.02391v3*, 2017.

P.Y. Simard, D. Steinkraus, and J.C. Platt. Best practices for convolutional neural networks applied to visual document analysis. *ICDAR*, pages 958–963, 2003.

A. Singh, L. Bergman, and S. Seregard. Uveal melanoma: epidemiologic aspects. *Clini. Opht. Onc.*, page 75–87, 2014.

T. Tartaglione, M. Pagliara, M. Sciandra, C. Caputo, R. Calandrelli, G. Fabrizi, S. Gaudino, M. Blasi, and C. Colosimo. Uveal melanoma: evaluation of extrascleral extension using thin-section mr of the eye with surface coils. *Radiol Med.*, 119(10):775–783, 2014.

N. Tustison, B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, and J. Gee. N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010.

S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J-C. Vilanova, L. Ramió-Torrentà, A. Rovira, A. Oliver, and X. Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach. *NeuroImage*, 2017.

B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. *CVPR*, pages 2921–2929, 2016.